

# Serotype cycles in cholera dynamics

Katia Koelle<sup>1,\*</sup>, Mercedes Pascual<sup>1</sup> and Md. Yunus<sup>2</sup>

<sup>1</sup>The Department of Ecology and Evolutionary Biology, 2019 Kraus Natural Science Building, University of Michigan, 830 North University Avenue, Ann Arbor, MI 48109-1048, USA

<sup>2</sup>International Centre for Diarrheal Disease Research, Dhaka 1000, Bangladesh

Interest in understanding strain diversity and its impact on disease dynamics has grown over the past decade. Theoretical disease models of several co-circulating strains indicate that incomplete cross-immunity generates conditions for strain-cycling behaviour at the population level. However, there have been no quantitative analyses of disease time-series that are clear examples of theoretically expected strain cycling. Here, we analyse a 40-year (1966–2005) cholera time-series from Bangladesh to determine whether patterns evident in these data are compatible with serotype-cycling behaviour. A mathematical two-serotype model is capable of explaining the oscillations in case patterns when cross-immunity between the two serotypes, Inaba and Ogawa, is high. Further support that cholera's serotype-cycling arises from population-level immunity patterns is provided by calculations of time-varying effective reproductive rates. These results shed light on historically observed serotype dominance shifts and have important implications for cholera early warning systems.

**Keywords:** cholera serotypes; strain cycling; cross-immunity; cholera two-strain model; disease time series

## 1. INTRODUCTION

Patterns of strain diversity and evolution may result from intrinsic characteristics of diseases, such as the persistence of infection and the interaction of strain variants (Grenfell *et al.* 2004). Interactions between strain variants include priming each other through antibody-dependent enhancement (Ferguson *et al.* 1999), directly competing within a host in times of co-infection, interfering with each other through generating a convalescent period in the host (Earn *et al.* 2000) or inducing cross-immunity (Andreasen *et al.* 1997). When strain interaction is through cross-immunity, herd-immunity levels generate selection pressures on strain evolution and impact disease dynamics. In the case of pathogens such as influenza, continual antigenic drift results from partial cross-immunity between sufficiently similar strains and the large number of possible variants. For pathogens with only a few possible strain variants, theoretical studies predict strain cycling as the predominant pattern (Andreasen *et al.* 1997; Gupta *et al.* 1998; Dawes & Gog 2002). However, due to the multiplicity of strains that underlies many diseases and the lack of sufficiently long disease datasets, a clear empirical example of this theoretical expectation is still lacking.

Here, we analyse a cholera dataset from Matlab, Bangladesh to determine whether the serotype patterns observed in the time-series provide such an example. The disease is caused by the Gram-negative bacterium *Vibrio cholerae*, and consists of two pathogenic serogroups, 01 and 0139. These serogroups differ significantly in their antigenic lipopolysaccharide characteristics, with no

apparent cross-immunity between each other (Mooi & Bik 1997). However, within the 01 serogroup classification, cholera cases are identified as being of a particular biotype (classical or El Tor) and of a particular serotype (Inaba or Ogawa, and very rarely Hikojima). Cholera biotypes are distinct phenotypes that differ with respect to the severity of their infections (and consequently their infection-to-case ratios), ability to survive outside the human host and seasonality patterns (Woodward & Mosley 1972; Glass *et al.* 1982; Reidl & Klose 2002; Koelle *et al.* 2005a). Serotypes differ from one another only with respect to antigenic determinants that are present on their O-antigen capsule. Inaba organisms have antigenic determinants identified as A and C, while Ogawa organisms have determinants A and B (Stroher *et al.* 1992). Differences in the serotypes' antigenic determinants are not known to confer differences in any phenotypic characteristics, such as the duration of the infectious period, the strength of the hosts' immune response or the recovery rate. However, these antigenic determinants play a critical role in inducing host immunity. Owing to the common presence of the 'A' antigenic determinant, cross-immunity between serotypes exists. Strong evidence for this cross-immunity comes from patients who, being infected with one serotype, undergo serological conversion against the heterologous serotype (Benenson *et al.* 1968). Even though present, cross-immunity is thought to be incomplete for several reasons: (i) rare within-host serotype conversions are occasionally associated with relapses of symptoms (Gangarosa *et al.* 1967); (ii) cholera patients infected with a recently appeared serotype are generally older than patients infected with the endemic serotype (ProMed Digest 2005); and (iii) founder serotypes are often replaced by the heterologous serotype in epidemic areas once considerable serotype-specific immunity has been generated in the population (Felsenfeld 1966).

\* Author for correspondence (kkoelle@umich.edu).

The electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2006.3668> or via <http://www.journals.royalsoc.ac.uk>.

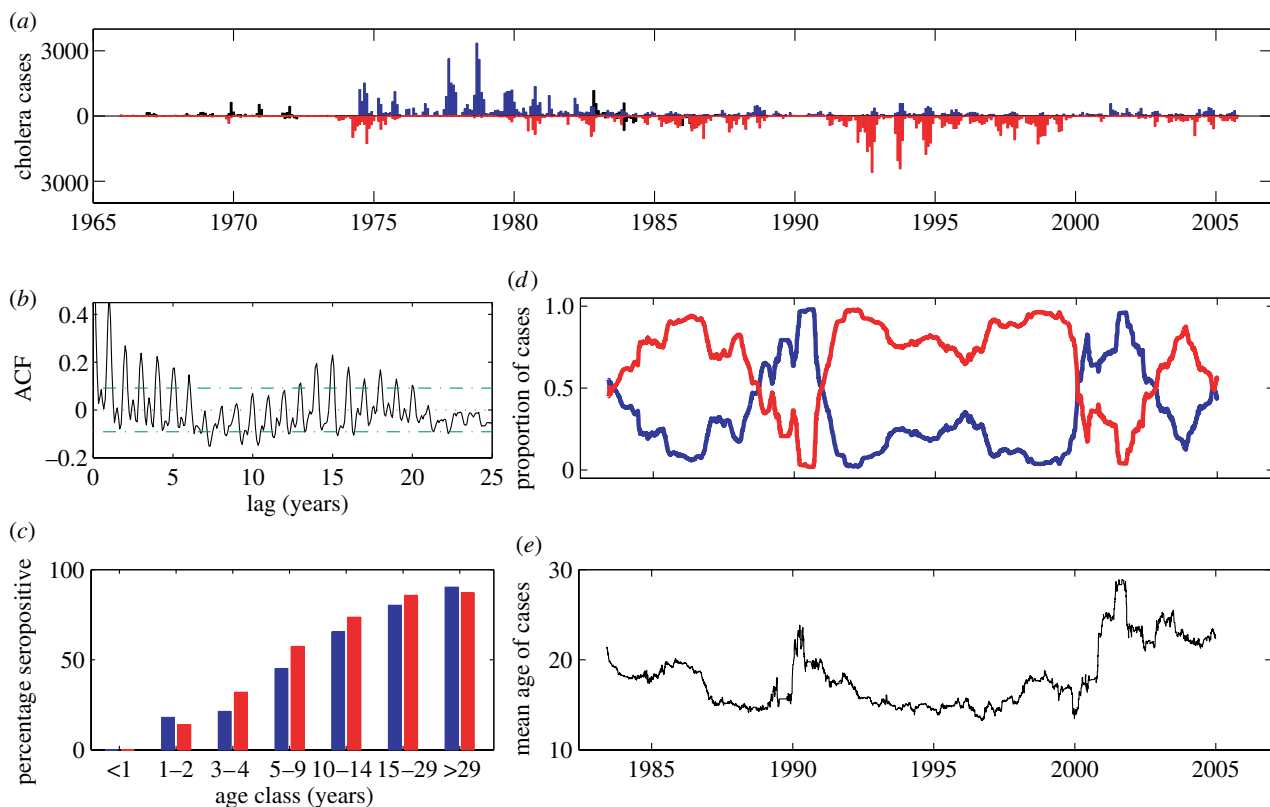


Figure 1. Observed cholera case data and age-incidence data. Inaba data are shown in blue; Ogawa data are shown in red. (a) Cholera case data over the period January 1966–October 2005 from Matlab, Bangladesh. Symptomatic cholera case data were obtained from the hospital records of the International Centre for Diarrheal Disease Research, Bangladesh (ICDDR,B), a demographic surveillance site located 55 km southeast of Dhaka. These symptomatic cholera cases were multiplied by their infection-to-case ratios ( $C_{ET}=15:1$  for El Tor cases and  $C_{CL}=3.7:1$  for classical cases) to yield the case data. Estimates of infection-to-case ratios were obtained from field studies in the Matlab area (Woodward *et al.* 1970; Woodward & Mosley 1972). Cases are separated into serotypes (Inaba and Ogawa) and biotypes (classical and El Tor). For clarity, classical cases are shown in black, with the axis on which they appear indicating their serotype. (b) Autocorrelation function of cholera case data. To show the significance of the 15-year period in cholera cases ( $p<0.05$ ), 95% confidence intervals are included (green, dashed). (c) Percentage of the Matlab population seropositive for Inaba or Ogawa, as defined by antibody titre of 1:20 or greater. The serological survey was conducted in 1965 on a randomized age-stratified sample population. (d) Proportion of symptomatic cases of Inaba and Ogawa serotypes over the period January 1983–October 2005, temporally aligned with figure 1a. (e) Mean age of symptomatic cholera cases over the same period (1983–2005). Both the proportions (3.1d) and the mean ages (3.1e) were calculated using a 1-year sliding window of cases. Age data were unavailable prior to 1983.

The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning *et al.* 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

## 2. DISEASE DATA

In areas where cholera is endemic, it has long been noticed that serotypes tend to fluctuate in dominance (Pollitzer 1959), with shifts occurring in the intervals between epidemics of the disease (Venkatraman 1947; Felsenfeld 1966). However, the reason for these fluctuations has not been understood, and the identification of regularities of these patterns is absent (Bhaskaran & Gorrill 1957; ProMed Digest 2005). A long-term cholera dataset provided by the International Centre for Diarrheal Disease Research, Bangladesh allows for the quantitative study of the origin of these serotype patterns (figure 1a). The cholera case data indicate a temporal separation of serotype dominance: while the Inaba serotype was dominant between the years 1966 and 1983, Ogawa has been the predominant serotype since that time.

The autocorrelation function (ACF) of cholera cases over the period 1966–2005 (figure 1b) indicates a statistically significant 15-year period between Inaba- and Ogawa-dominant case peaks ( $p<0.05$ ). Cholera seasonality is also evident in the ACF, generating yearly peaks in the correlation coefficient. An age-stratified serological study conducted in this surveillance area in the summer of 1965 provides additional serotype-specific information (figure 1c; Mosley *et al.* 1968); the percentage of sampled individuals who tested positive for Ogawa vibriocidal antibodies was greater than that of the individuals who tested positive for Inaba antibodies in age classes 3–4 to 15–29 year olds. These results indicate that Ogawa was the predominant serotype for a period before 1966, and that the transition in the early 1980s from Inaba to Ogawa is not part of a long-term trend.

To supplement the cholera case data, we computed the proportion of Inaba and Ogawa cases over the period 1983–2005 (figure 1d) and plotted the mean age of infected individuals over this time (figure 1e). These plots show that changes in serotype dominance are associated with increases in the mean age of cholera cases. Most strikingly, we see peaks in the mean age of cases in the

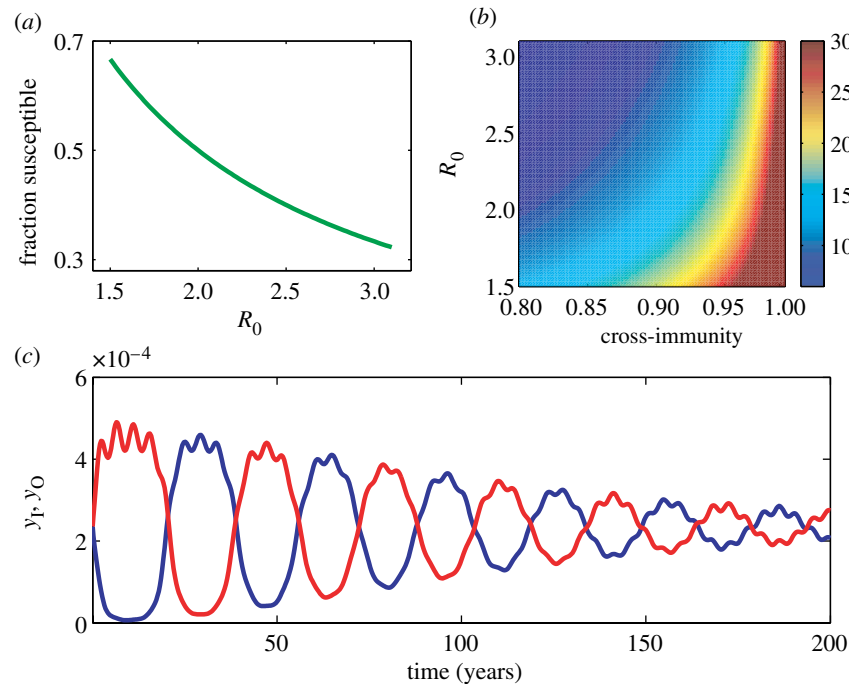


Figure 2. Analysis of the two-serotype cross-immunity model, given by equations (3.1a–e). (a) The equilibrium fraction susceptible to becoming infectious,  $s^*$ , as a function of the basic reproductive rate,  $R_0$ . (b) Sensitivity of the inter-epidemic period to  $R_0$  and  $\gamma$ . Values for  $s^*$  and the inter-epidemic period were analytically computed from the eigenvalues of the linearized system (Kamo & Sasaki 2002). The range for cholera's  $R_0$  was computed with parameters,  $G = 34.28$  years (Mostafa *et al.* 1996) and  $A = 11.2$  years, yielding  $R_0 = 1.53$ – $3.06$ . The average age of first infection,  $A$ , was estimated by applying Muench's catalytic model (Muench 1959) to the 1965 serological survey data (figure 1c). Note that the average age of first infection differs from the average age of infection (figure 1e), because individuals may become infected at a later point in life with the heterologous serotype. Independently estimated parameters in the model include the duration of infectiousness  $1/\sigma = 9.5$  days (Lindenbaum *et al.* 1967; Woodward & Mosley 1972) and the population turnover rate  $\mu = 1/38.8$  year $^{-1}$  (which includes turnover from both emigration and mortality; Mostafa *et al.* 1996). (c) Simulated fluctuations in proportion of Inaba and Ogawa cases, shown over a period of 200 years. These dynamics show an oscillatory approach to equilibrium,  $y_I^*$  and  $y_O^*$ , with two frequencies present: a 15-year asynchronous inter-epidemic period and a less dominant 4.5-year synchronous period. Dynamics were simulated with  $R_0 = 2.9$  and  $\gamma = 0.956$ .

early 1990s and in 2002/2003 when Inaba cases, although few in absolute number, are highest in proportion. A less striking, but nevertheless important, pattern is the slow decrease in mean age of cases in the early 1980s, as the Ogawa serotype becomes established. The association between serotype shifts and age incidence patterns seen here is in agreement with the previous observations, for example, by medical personnel in nearby Calcutta, India (ProMed Digest 2005).

In addition to the observed cholera data shown in figure 1, we have detailed demographic data over this period, with monthly birth and death information and decadal census data (not shown). The size of the population in the surveillance area is growing at a rate of 0.77% per year, with a current population size of 225 000.

### 3. THE MATHEMATICAL MODEL

We turn to the hypothesis that the observed serotype dynamics arise from the immunological interaction between the two serotypes, and thereby constitute part of a strain cycle. To first determine whether strain cycling is theoretically expected for a two-serotype system such as cholera, and to be able to quantify characteristics of these dynamics, we revisit a published two-strain model by Kamo & Sasaki (2002). Within this mathematical framework, cross-immunity acts by reducing the ability of a host to become infected with the heterologous strain. An

individual who has been previously infected can therefore still become infected with the heterologous serotype, but at a reduced rate. Alternative definitions and models for cross-immunity exist (e.g. Andreasen *et al.* 1997; Gupta *et al.* 1998; Gog & Swinton 2002); we use this representation because volunteer studies have shown that patients re-challenged with the heterologous serotype of cholera rarely become infected. When symptoms do occur, they are milder and *V. cholerae* excretion levels are reduced (Cash *et al.* 1974), arguing for a concomitant reduction in transmissibility, which is not incorporated in the model framework we use here. The empirical evidence for the type of cross-immunity between serotypes therefore does not unequivocally support a particular model, e.g. Kamo & Sasaki (2002) versus Gupta *et al.* (1998), but a combination of models. The particular choice of model, however, does not affect our general conclusion (results not shown). Here, we use the model presented by Kamo & Sasaki (2002) owing to its analytical results. From Kamo & Sasaki (2002), we have

$$dy_I/dt = \beta[x_N + (1 - \gamma)x_0]y_I - (\sigma + \mu)y_I, \quad (3.1a)$$

$$dy_O/dt = \beta[x_N + (1 - \gamma)x_1]y_O - (\sigma + \mu)y_O, \quad (3.1b)$$

$$dx_N/dt = -\beta(y_I + y_O)x_N + \mu(1 - x_N), \quad (3.1c)$$

$$dx_I/dt = \beta x_N y_I - \beta(1 - \gamma)x_I y_O - \mu x_I, \quad (3.1d)$$

$$dx_O/dt = \beta x_N y_O - \beta(1 - \gamma)x_0 y_I - \mu x_O \quad \text{and}, \quad (3.1e)$$

$$dx_B/dt = \beta(1 - \gamma)x_1 y_O + \beta(1 - \gamma)x_0 y_I - \mu x_B. \quad (3.1f)$$

Here,  $y_I$  is the proportion of individuals infected with the Inaba serotype,  $y_O$ , the proportion of individuals infected with the Ogawa serotype,  $x_N$ , the proportion of hosts naive to both the serotypes,  $x_I$ , the proportion of hosts who have or have had an Inaba infection,  $x_O$ , the proportion of hosts who have or have had an Ogawa infection,  $x_B$ , the proportion of hosts who have or have had both Inaba and Ogawa infections. The present or previous infection with a serotype provides permanent immunity against re-infection with that serotype. Since  $x_N + x_I + x_O + x_B = 1$ , we use only the first five equations (3.1a–e) to completely describe the system. The parameters in the model are duration of infectiousness,  $1/\sigma$ , population turnover rate,  $\mu$ , transmissibility,  $\beta$ , and degree of cross-immunity,  $\gamma$ .

To see whether this model can generate the cholera patterns of figure 1, we first estimated parameter values for  $\sigma$  and  $\mu$  from Matlab's demographic and cholera literature (figure 2). Two unknown parameters remained: the transmission rate,  $\beta$ , and the degree of cross-immunity,  $\gamma$ . Estimates of  $\beta$  can be computed from the basic reproductive rate,  $R_0$ , by  $\beta = R_0(\sigma + \mu)$ . ( $R_0$  for a two-strain system is defined as the number of secondary cases one infected individual begets in a population that is completely susceptible to both strains.)  $R_0$  can be estimated by the equation  $G/A$  for a single endemic strain in a growing population (May & Anderson 1985), with  $G$  being the reciprocal of the birth rate, and  $A$ , the average age of first infection. This equation naturally also holds for a two-strain system with complete cross-immunity. For a two strain-system in which there is no cross-immunity,  $R_0$  is estimated by  $G/2A$  (Gupta *et al.* 1994). For cholera, because some cross-immunity between the serotypes exists, we (for now) simply bound  $R_0$  by these upper and lower limits (figure 2).

The degree of cross-immunity is a less easily identifiable parameter, because it is not consistently defined in the literature and owing to the limited number of volunteer re-challenge experiments and field studies. However, the studies that have taken place indicate that serotype cross-immunity is almost complete (Levine 1978; Levine *et al.* 1979). Additional patient studies show that a significant rise in vibriocidal antibodies to the heterologous serotype occurs in approximately 90% of infections (Benenson *et al.* 1968), further supporting a high level of cross-immunity, since vibriocidal levels are indicative of protection against cholera infection (Mosley 1969). Although considered high, cross-immunity is thought to be incomplete for the reasons previously mentioned (within-host serotype conversions, age-incidence patterns and serotype replacement patterns in epidemic areas).

Given our estimates of  $\sigma$  and  $\mu$ , we first look at the sensitivity of the model's behaviour as a function of the lesser known parameters,  $R_0$  (proportional to  $\beta$ ) and cross-immunity,  $\gamma$ . A mathematical analysis of this two-strain model (Kamo & Sasaki 2002) showed that this system has three fixed points. Two are the simple cases of one serotype being present and the other serotype being absent. The endemic fixed-point is the only stable fixed point when  $R_0 > 1$ , with solutions given by Kamo & Sasaki (2002). At this stable fixed point, we can define  $s_I^*$  as the equilibrium fraction of hosts susceptible to becoming infected with the Inaba serotype; mathematically,  $s_I^* = x_N^* + (1 - \gamma)x_O^*$ , as is evident from equation (3.1a).

Similarly, the proportion susceptible to becoming infected with Ogawa is  $s_O^* = x_N^* + (1 - \gamma)x_I^*$ . Since the serotypes Inaba and Ogawa are symmetrical in their parameter values,  $s_I^* = s_O^*$ ; let  $s^*$  equal this value. From equations (3.1a,b), we find that  $s^* = 1/R_0$  and thereby independent of the degree of cross-immunity,  $\gamma$ . In figure 2a, we plot these analytically computed values for  $s^*$  as a function of  $R_0$ . We can now use the serological data presented in figure 1c along with age-structured demographic data (Mosley *et al.* 1968) to compute cholera's  $s_I^*$  and  $s_O^*$  in 1965:  $s_I^* = 0.36$  and  $s_O^* = 0.32$ . Taking  $s^*$  to be an average of these values, our estimate of  $R_0$  can be approximated as  $R_0 = 2.9$ . This value is on the high end of our  $R_0$  range and is reasonable in light of the expected high degree of cross-immunity seen from volunteer re-challenge studies.

Through a linearization of the system and an analysis of the resulting eigenvalues, Kamo and Sasaki also identified two natural frequencies present at this fixed point,  $\omega_1$  and  $\omega_2$ ;  $\omega_1$  corresponded to the natural frequency of the synchronized dynamics of the two strains, while  $\omega_2$  corresponded to the natural frequency of the anti-phase dynamics of the two strains. The anti-phase natural frequency,  $\omega_2$ , and its associated period,  $2\pi/\omega_2$ , can be computed directly from the imaginary part of the relevant eigenvalues. Since the fluctuations of the strains associated with this natural frequency are anti-phase (Kamo & Sasaki 2002), the period between one strain's peak and the other strain's peak, i.e. the inter-epidemic period, is simply  $\pi/\omega_2$ . We plot this inter-epidemic period of the damped oscillations in figure 2b, as a function of  $R_0$  and the degree of cross-immunity,  $\gamma$ . For a given value of  $R_0$ , increasingly higher cross-immunity levels lead to serotype cycles of longer periods. As expected, when cross-immunity between strains is complete ( $\gamma = 1$ ), the cycle length is infinite, effectively resulting in the competitive exclusion of one strain. Given the value of  $R_0$  estimated from our calculated  $s^*$  in figure 2a, the 15-year period of oscillation observed in our serotype dynamics (figure 1b) is consistent with the strain dynamics of the model when the degree of cross-immunity value is  $\gamma = 0.956$ .

We simulated the model given by equations (3.1a–e) in figure 2c with our refined parameter values for  $\gamma$  and  $R_0$ . As expected, the transient dynamics show the 15-year anti-phase behaviour for the proportion of infected individuals,  $y_I$  and  $y_O$ . Also visible in this simulation is another frequency, associated with a shorter period. This frequency ( $\omega_1$ ) corresponds to the previously mentioned synchronous fluctuations of the two strains, with a period of 4.5 years, calculated directly from the relevant eigenvalues. This period is not evident in cholera's ACF (figure 1b). This may be because the second natural frequency mode, corresponding to  $\omega_2$ , contributes more to the amplitude response of the linearized system when perturbed by demographic or environmental noise. Indeed, when we calculate the friction coefficients of the two modes (Kamo & Sasaki 2002), we find that  $|\nu_2| \ll |\nu_1|$ .

Seasonal fluctuations in the transmission rate,  $\beta$ , have been shown capable of generating interannual disease cycles in single-strain deterministic models (Schwartz & Smith 1983; Schwartz 1992). Likewise, Kamo & Sasaki (2002) showed that when seasonality is included in simulations of



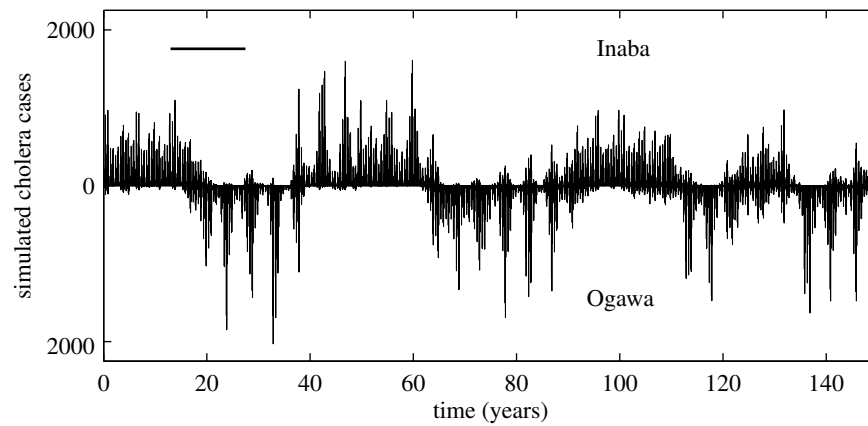


Figure 3. Stochastic realization of the two-serotype mathematical model, with seasonality, shown over a period of 150 years. The first 500 years of simulated data were removed as transients. The horizontal black bar measures 15 years. Simulated cases were aggregated in monthly intervals to allow for a comparison with figure 1a. The parameters,  $\mu$ ,  $\sigma$ ,  $R_0$  and  $\gamma$  are the same as in figure 2c. Population size for this simulation was one-tenth the size of Bangladesh, and cholera case numbers were adjusted to Matlab's population size of 225 000 prior to plotting. The model was simulated for an effectively larger population size than Matlab alone, such that stochastic extinctions were rare. These simulations addressed whether the decaying cycles of the deterministic system become persistent in the presence of noise and seasonality, without introducing modifications and parameters to the original serotype model to account for strain re-introductions after extinction (see §5).

this two-strain model, the system can sustain interannual cycles of various periods, potentially a result of resonance with the natural frequencies of oscillation. Since cholera is highly seasonal (Glass *et al.* 1982), we simulated the deterministic cholera two-serotype system with realistic seasonality patterns derived from empirical monthly case distributions (electronic supplementary material). However, in our system, seasonality alone simply generated annual cycles with no interannual cycles evident (results not shown). Since the model we use is one that has been shown capable of generating interannual cycles when deterministic seasonal forcing is present (Kamo & Sasaki 2002), the inability of seasonal forcing to sustain these cycles in our system must be due to cholera's specific parameter values. We suspect that either the very high degree of cross-immunity between the serotypes or the specific seasonal forcing function of cholera does not generate sufficient resonance with the natural oscillation frequencies to sustain interannual cycles in a purely deterministic framework.

It is also well known that demographic noise can sustain oscillations that would otherwise decay in time (Bartlett 1957; Bailey 1975). We therefore simulated a stochastic event-driven implementation of the two-serotype model. Simulations both with and without seasonality in transmission rates sustained the oscillations in case numbers (results not shown for without-seasonality case and figure 3 for with-seasonality case), with long inter-epidemic periods of approximately 15 years. Variability in the duration of serotype dominance was present, with the inter-epidemic period ranging between 10 and 20 years.

#### 4. EFFECTIVE REPRODUCTIVE RATES

If we assume that this cross-immunity model adequately describes the interaction between the cholera serotypes, we can use Matlab's demographic data and observed case data to reconstruct time-varying effective reproductive rates (figure 4), an indicator of serotype fitness levels in the population. The effective reproductive rate over time for Inaba is given by  $R_I(t) = (\beta[x_N(t) + (1 - \gamma)x_O(t)])/(\sigma + \mu)$ ,

and similarly for Ogawa,  $R_O(t) = (\beta[x_N(t) + (1 - \gamma)x_I(t)])/(\sigma + \mu)$ . With known parameters  $\beta$ ,  $\sigma$ ,  $\mu$  and  $\gamma$ , and initial values for  $x_N$ ,  $x_O$  and  $x_I$  that are consistent with the serological data (figure 1c), these time-varying effective reproductive rates can be reconstructed through an iterative approach (electronic supplementary material). The relative values of the effective reproductive rates quantify the magnitude of selection against the less-fit serotype (figure 4b), and can be computed by standardizing the serotypes' effective reproductive rates to the one with higher fitness. Relative fitness levels are thereby computed as  $r_I(t) = (R_I(t)/\max(R_I(t), R_O(t)))$  for Inaba and  $r_O(t) = (R_O(t)/\max(R_I(t), R_O(t)))$  for Ogawa.

Effective reproductive rates,  $R_I(t)$  and  $R_O(t)$ , fluctuate in response to changes in immunity levels, providing evidence for immunity-driven cholera case dynamics. Note that the number of Inaba cases generally grows until 1979 when the reconstructed effective reproductive values become less than 1 (for both serotypes), and likewise that the number of Ogawa cases grows during the period 1990–1993, then decreases in the rest of the 1990s and stabilizes at very low levels until the present time, in agreement with the dynamics expected from the effective reproductive rates.

Changes in the relative magnitude of the effective reproductive rates arise as a response to the population dynamics of the cholera cases; Ogawa becomes the fitter serotype after the long period in the 1970s of predominantly Inaba cases, while Inaba becomes the fitter serotype again in the early 1990s, after a period of predominantly Ogawa cases. This observation indicates that partial cross-immunity between Inaba and Ogawa serotypes generates conditions for frequency-dependent selection pressures. Moreover, switches in the identity of the fitter serotype (figure 4b) are seen to occur near the peaks of the outbreaks in 1978 and 1993 (figure 1a), with the highest relative fitness differences occurring when effective reproductive rates are near their minima, in 1983 and 2000. Given that cases are declining when effective reproductive rates are less than 1, a subsequent increase in cases when these rates

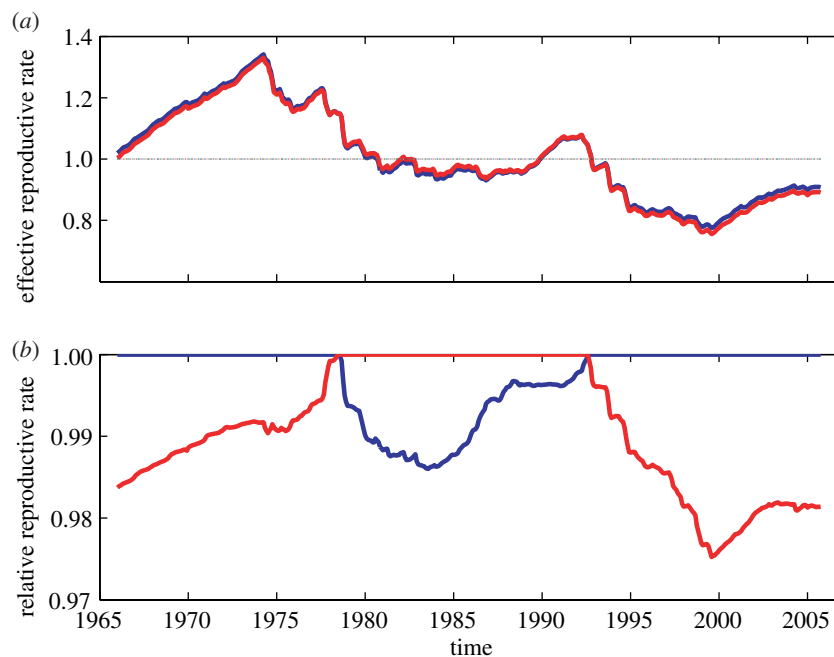


Figure 4. Reconstructed effective reproductive rates. (a) Time-varying effective reproductive rates for Inaba and Ogawa. (b) Time-varying relative effective reproductive rates. Both El Tor and classical cases were included in the reconstruction of reproductive rates, with full immunity between classical and El Tor biotypes of a given serotype. Initial conditions for the reconstruction were  $x_N = 0.33$ ,  $x_I = 0.19$ ,  $x_O = 0.32$  and  $x_B = 0.15$ , consistent with the average fraction susceptible to becoming infectious,  $s^*$ , of 0.34, and consistent with the serological evidence that the fraction susceptible to Inaba infection is greater than that susceptible to Ogawa infection in 1965 (figure 1c).

become greater than 1 are likely to occur predominantly with the fitter serotype. Hence, fluctuations in relative effective reproductive rates can explain the previously noted pattern of serotype re-emergences occurring in the time-intervals between epidemic outbreaks (Venkatraman 1947).

## 5. DISCUSSION

Our analysis investigated whether patterns in cholera's serotype fluctuations provided quantitative empirical evidence for theoretically expected strain-cycling behaviour. A mathematical model was able to capture the observed cyclical serotype dynamics seen in the cholera data when all four of its parameter values were either independently estimated from demographic and epidemiological data ( $\mu$  and  $\sigma$ ) or had reasonable ranges independently estimated ( $R_0$  and  $\gamma$ ). Combined with the reconstructed effective reproductive rates, these analyses indicate that high, but incomplete, cross-immunity between cholera's serotypes is a well-supported explanation for the observed long-term changes in serotype dominance. This research provides insight into the reason for serotype dominance changes in cholera; it also helps us understand how epidemiological factors (such as  $R_0$  and the degree of cross-immunity) affect the period of these serotype cycles. Moreover, the two-serotype model with incomplete cross-immunity provides an explanation for the increase in the mean age of infected individuals, which is observed when one cholera serotype re-emerges (ProMed Digest 2005), a pattern we also see in our data (figure 1e,f). When a serotype re-emerges, older individuals who are only immune to the previously dominant serotype may come in contact with the re-emerging serotype, thereby becoming infected and raising the mean age of cholera cases.

Previous research of ours using nonlinear time-series models to fit cholera incidence data has shown that instead of conferring permanent immunity to a previously infected host, cholera may confer temporary immunity of the order of about a decade in duration (Koelle & Pascual 2004; Koelle *et al.* 2005b). These analyses did not consider cholera cases at the level of the specific serotypes, and may therefore reflect phenomenological waning of immunity caused by serotype shifts at the level of the population. However, if we consider that cholera of a given serotype may induce less than lifelong immunity, we would expect serotype cycles to be faster than cycles generated with the assumption of permanent immunity, for given values of  $R_0$  and cross-immunity. Temporary immunity is still capable of generating serotype cycles of long periods when cross-immunity levels between the two serotypes is sufficiently high (results not shown).

Insights gained from considering cholera dynamics at the serotype level also have predictive value, suggesting general qualitative trends. Given the calculated trajectories in the effective reproductive rates, we would expect the coming decade to show an increase in absolute case numbers, with Inaba re-emerging as the predominant serotype. In the early period of Inaba re-emerging, adults should also become infected, but the mean age of cholera patients should decrease thereafter as Inaba becomes increasingly established in the population. More quantitative predictions require further developments. Of particular relevance is the demonstrated high sensitivity of the model dynamics to the cross-immunity between Inaba and Ogawa. It is therefore critical to obtain better estimates of this parameter for predicting the timing of serotype switching. Complementary approaches include statistical modelling of cholera dynamics based on this serotype model, along with stronger surveillance programs

and more comprehensive re-challenge studies. The extreme sensitivity to the cross-immunity parameter also some of the variability in immunity cycle lengths of the stochastic model (figure 3), and therefore why trends in serotype dominance may seem irregular or not readily apparent (Bhaskaran & Gorrill 1957; ProMed Digest 2005).

Demographic stochasticity was considered here for a much larger population than the one for the surveillance region. The scope of the stochastic simulations was to verify that in the presence of noise, the decaying oscillations of the deterministic model become persistent. For small populations, extinctions are inevitable. While these extinctions are realistic, they require the explicit treatment in the models of additional processes associated with strain re-introductions or 'rescue' effects. In the case of cholera, the most important process to be considered is primary transmission from an environmental, aquatic reservoir (Hartley *et al.* 2006), a parameter for which independent estimates are not currently available. Our model formulation with only a direct transmission pathway should now be extended to include primary transmission, and fitted to the data in the presence of stochasticity, to consider smaller population sizes, evaluate the different transmission pathways and the predictability of the dynamics.

Mathematical models for cholera early warning systems will also need to examine how to couple the dynamics of serotype-specific immunity levels to variable transmission rates driven by environmental factors. Extrinsic factors, such as the El Niño-Southern Oscillation (ENSO) and long-term variability in northeast India rainfall, have previously been shown to play a role in the interannual variability of cholera at shorter periods (Pascual *et al.* 2000; Koelle *et al.* 2005b). Although the two-serotype model demonstrates that general trends in cholera cases and serotype fluctuations can be explained through intrinsic immunity mechanisms, several pieces of evidence also point here towards extrinsic factors as additional players in driving disease dynamics. In this model, variability in the reconstructed effective reproductive rates was generated solely by fluctuations in population-level susceptibility levels (figure 4), while the transmission rate,  $\beta$ , was assumed a constant. It is this transmission term that is modulated by environmental drivers. For example, the resurgence of cases in the time-interval 1982–1984 by the classical biotype (of both Inaba and Ogawa serotypes) occurs when effective reproductive rates, only influenced by immunity levels, are negative. This period follows the strong 1982–1983 El Niño event, which may have transiently raised the transmission rate of cholera, thereby effectively yielding overall effective reproductive rate greater than 1 for the period 1982–1984. Similarly, the negative ENSO anomalies present in the early 1970s, indicative of La Niña conditions, may have transiently lowered cholera transmission rates, thereby explaining the low number of cholera cases during this time, despite the high effective reproductive rates shown in figure 4. How environmental forcing at characteristic time-scales influences the period of the serotype cycles remains to be addressed more fully in future research. The variability observed in the stochastic simulations around a mean period cycle of 15 years suggests that cycle length may indeed vary, although these long-term cycles are generally expected to be sustained.

Major advances in understanding and predicting disease dynamics have been made for childhood diseases, such as measles (Earn *et al.* 2000; Finkenstädt & Grenfell 2000), where immunity is permanent and strain transcendent. However, among the most ubiquitous diseases are influenza, malaria and diarrhoeal diseases, many of which have underlying strain structure with only partial cross-immunity between strains. With advances in molecular epidemiology and phylogenetic approaches, the dynamics of these multi-strain diseases are starting to be examined, with implications for early warning systems and vaccine development. Here, we have demonstrated that high cross-immunity between cholera's serotypes is a well-supported mechanism that sets these strains in motion.

We thank Kim Streatfield at the ICDDR,B for support with the demographic data and the NCEAS Seasonality of Infectious Diseases working group, in particular Ottar Bjornstad, David Alonso and Bruce Kendall, for helpful comments, and the Centre for the Study of Complex Systems at UM for cluster access. K.K. thanks the Rackham Graduate School for travel funds and G. S. S. for technical advice and feedback. This work was supported by EEB graduate fellowship support to K.K. and by the joint support of the NSF-NIH (Ecology of Infectious Diseases) and NOAA (Oceans and Health) to M.P.

## REFERENCES

- Andreasen, V., Lin, J. & Levin, S. A. 1997 The dynamics of co-circulating influenza strains conferring partial cross-immunity. *J. Math. Biol.* **35**, 825–842. (doi:10.1007/s002850050079)
- Bailey, N. T. J. 1975 *The mathematical theory of infectious diseases and its applications*. London, UK: Charles Griffin & Company Ltd.
- Bartlett, M. S. 1957 Measles periodicity and community size. *J. R. Stat. Soc. Ser. A* **120**, 48–70.
- Benenson, A. S., Saad, A. & Mosley, W. H. 1968 Serological studies in cholera. 2. The vibriocidal antibody response of cholera patients determined by a microtechnique. *Bull. World Health Organ.* **38**, 277–285.
- Bhaskaran, K. & Gorrill, R. H. 1957 A study of antigenic variation in *Vibrio cholerae*. *J. Gen. Microbiol.* **16**, 721–729.
- Cash, R. A., Music, S. I., Libonati, J. P., Craig, J. P., Pierce, N. F. & Hornick, R. B. 1974 Response of man to infection with *Vibrio cholerae*. II. Protection from illness afforded by previous disease and vaccine. *J. Infect. Dis.* **130**, 325–333.
- Dawes, J. H. P. & Gog, J. R. 2002 The onset of oscillatory dynamics in models of multiple disease strains. *J. Math. Biol.* **45**, 471–510. (doi:10.1007/s00285-002-0163-9)
- Earn, D. J., Rohani, P., Bolker, B. M. & Grenfell, B. T. 2000 A simple model for complex dynamical transitions in epidemics. *Science* **287**, 667–670. (doi:10.1126/science.287.5453.667)
- Felsenfeld, O. 1966 A review of recent trends in cholera research and control. *Bull. World Health Organ.* **34**, 161–195.
- Ferguson, N., Anderson, R. & Gupta, S. 1999 The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens. *Proc. Natl Acad. Sci.* **96**, 790–794. (doi:10.1073/pnas.96.2.790)
- Finkenstädt, B. F. & Grenfell, B. T. 2000 Time series modelling of childhood diseases: a dynamical systems approach. *Appl. Stat.* **49**, 187–205.

- Gangarosa, E. J., Sanati, A., Saghari, H. & Feeley, J. C. 1967 Multiple serotypes of *Vibrio cholerae* isolated from a case of cholera. *Lancet* **1**, 646–648. (doi:10.1016/S0140-6736(67)92542-1)
- Glass, R. I., Becker, S., Huq, M. I., Stoll, B. J., Khan, M. U., Merson, M. H., Lee, J. V. & Black, R. E. 1982 Endemic cholera in rural Bangladesh, 1966–1980. *Am. J. Epidemiol.* **116**, 959–970.
- Gog, J. R. & Swinton, J. 2002 A status-based approach to multiple strain dynamics. *J. Math. Biol.* **44**, 169–184. (doi:10.1007/s002850100120)
- Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A. & Holmes, E. C. 2004 Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–332. (doi:10.1126/science.1090727)
- Gupta, S., Trenholme, K., Anderson, R. M. & Day, K. P. 1994 Antigenic diversity and the transmission dynamics of *Plasmodium falciparum*. *Science* **263**, 961–963.
- Gupta, S., Ferguson, N. & Anderson, R. 1998 Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents. *Science* **280**, 912–915. (doi:10.1126/science.280.5365.912)
- Hartley, D. M., Morris, J. G. & Smith, D. L. 2006 Hyperinfectivity: a critical element in the ability of *V. cholerae* to cause epidemics? *PLoS Med.* **3**, 63–69. (doi:10.1371/journal.pmed.0030007)
- Kamo, M. & Sasaki, A. 2002 The effect of cross-immunity and seasonal forcing in a multi-strain epidemic model. *Physica D* **165**, 228–241. (doi:10.1016/S0167-2789(02)00389-5)
- Koelle, K. & Pascual, M. 2004 Disentangling extrinsic from intrinsic factors in disease dynamics: a nonlinear time series approach with an application to cholera. *Am. Nat.* **163**, 901–913. (doi:10.1086/420798)
- Koelle, K., Pascual, M. & Yunus, M. 2005a Pathogen adaptation to seasonal forcing and climate change. *Proc. R. Soc. B* **272**, 971–977. (doi:10.1098/rspb.2004.3043)
- Koelle, K., Rodó, X., Pascual, M., Yunus, M. & Mostafa, G. 2005b Refractory periods and climate forcing in cholera dynamics. *Nature* **436**, 696–700. (doi:10.1038/nature03820)
- Levine M. M. 1978 Immunity to cholera as evaluated in volunteers. In: *Cholera and related diarrheas, 43rd Nobel Symposium, Stockholm*, pp. 195–203. Karger, Basel.
- Levine, M. M. *et al.* 1979 Immunity of cholera in man: relative role of antibacterial versus antitoxic immunity. *Tran. R. Soc. Trop. Med. Hyg.* **73**, 3–9. (doi:10.1016/0035-9203(79)90119-6)
- Lindenbaum, J., Greenough, W. R. & Islam, M. R. 1967 Antibiotic therapy of cholera. *Bull. World Health Organ.* **36**, 871–883.
- Manning, P. A., Stroehrer, U. H. & Morona, R. 1994 Molecular basis of O-antigen biosynthesis in *Vibrio cholerae* 01: Ogawa–Inaba switching. In *Vibrio cholerae and cholera: molecular to global perspectives* (ed. I. K. Wachsmuth, P. A. Blake & O. Olsvik), pp. 77–94. Washington, DC: American Society of Microbiology.
- May, R. M. & Anderson, R. M. 1985 Endemic infections in growing populations. *Math. Biosci.* **77**, 141–156. (doi:10.1016/0025-5564(85)90093-8)
- Mooi, F. R. & Bik, E. M. 1997 The evolution of epidemic *Vibrio cholerae* strains. *Trends Microbiol.* **5**, 161–165. (doi:10.1016/S0966-842X(96)10086-X)
- Mosley, W. H. 1969 The role of immunity in cholera: a review of epidemiological and serological studies. *Texas Rep. Biol. Med.* **27**, 227–241.
- Mosley, W. H., Benenson, A. S. & Barui, R. 1968 A serological survey for cholera antibodies in rural East Pakistan. *Bull. World Health Organ.* **38**, 327–334.
- Mostafa G., Ahmed K., Shaikh M. A. K., van Ginneken J. K. & Sarder A. M. 1996 *Demographic Surveillance System—Matlab. Registration of demographic events—1995*. International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh.
- Muench, H. 1959 *Catalytic models in epidemiology*. Cambridge, MA: Harvard University Press.
- Pascual, M., Rodó, X., Ellner, S. P., Colwell, R. & Bouma, M. J. 2000 Cholera dynamics and El Niño–southern oscillation. *Science* **289**, 1766–1769. (doi:10.1126/science.289.5485.1766)
- Pollitzer, R. 1959 Epidemiology. In *Cholera* (ed. World-Health-Organization), pp. 836–841. Geneva, Switzerland: World Health Organization.
- ProMed Digest, 2005 Cholera, diarrhea, and dysentery update, 2005. Vol. 22.
- Reidl, J. & Klose, K. E. 2002 *Vibrio cholerae* and cholera: out of the water and into the host. *FEMS Microbiol. Rev.* **26**, 125–139. (doi:10.1111/j.1574-6976.2002.tb00605.x)
- Schwartz, I. B. 1992 Small amplitude, long period outbreaks in seasonally driven epidemics. *J. Math. Biol.* **30**, 473–491. (doi:10.1007/BF00160532)
- Schwartz, I. B. & Smith, H. L. 1983 Infinite subharmonic bifurcations in an SEIR model. *J. Math. Biol.* **18**, 233–253. (doi:10.1007/BF00276090)
- Stroehrer, U. H., Karageorgos, L. E., Morona, R. & Manning, P. A. 1992 Serotype conversion in *Vibrio cholerae* 01. *Proc. Natl Acad. Sci.* **89**, 2566–2570. (doi:10.1073/pnas.89.7.2566)
- Venkatraman K. V. 1947 A note on the reappearance of the Inaba type of *V. cholerae* in the Madras province. In: *Report. Indian Research Fund Association* (ed. Board SA), p. 10, New Delhi.
- Woodward, W. E. & Mosley, W. H. 1972 The spectrum of cholera in rural Bangladesh. II. Comparison of El Tor Ogawa and Classical Inaba infection. *Am. J. Epidemiol.* **96**, 342–351.
- Woodward, W. E., Mosley, W. H. & McCormack, W. M. 1970 The spectrum of cholera in rural East Pakistan. I. Correlation of bacteriologic and serologic results. *J. Infect. Dis.* **121**, S10–S16.